

2000年東京版総合社会調査における標本の代表性

松 本 康*

1. 目 的
2. 方 法
3. 性別構成比の検討
4. 年齢別構成比の検討
5. 性別・年齢別構成比の検討
6. 要約と結論

要 約

2000年9月に実施した東京版総合社会調査の標本の代表性を、2000年10月に実施された国勢調査データとの比較によって検討した。比較は、性別、年齢、性別×年齢の3つの観点から行い、TGSS標本が、5つの調査地点のそれぞれと、5地点合計および東京都全体をどの程度代表しているのかを検討した。

その結果、年齢次元においては、標本抽出誤差の範囲を超えて、概して男性が過小に、女性が過大に代表されており、その要因は女性よりも男性の回答率が低かったためであると考えられた。また、年齢次元においては、標本抽出誤差を超えて、概して、若年層が過小に、高齢層が過大に代表されており、その要因もまた年齢が若ければ若いほど回答率が低かったためと考えられた。そして年齢によるバイアスは女性よりも男性に強く表れていた。ただし、5地点全体と東京都全体とのずれはわずかなものであり、TGSS標本を5地点を代表するものとみなす場合のバイアスと、東京都全体を代表するものとみなす場合のバイアスは、ほとんど同じであることがわかった。これらのバイアスは、必ずしも本調査に特有のものではなく一般に見られるものである。

バイアスを補正する方法としては、重みづけをするよりも、クロスセクションの分析に徹することが適切であると示唆した。その理由は、われわれの調査の目的が意見の分布を記述することではなく、変数間の関連を分析することであるからであり、また、重みづけによって社会経済的地位のような未知のバイアスを強めてしまう恐れがあるからである。

1. 目 的

2000年9月に実施した東京版総合社会調査

(TGSS)の標本は、理論的には20歳以上70歳未満の都民全体を代表するように設計されていた。したがって、調査の結果得られた標本は、都民全体の縮図であることが望ましい。この代表性が、

*東京都立大学大学院都市科学研究科

実際にどの程度確保されているのか、また確保されていないとすればどのようなバイアスがあるのかを検討することが、本稿の目的である。

すでに報告したように、本調査の標本は次のような手続きにしたがって抽出された（詳しくは松本・原田 2001参照）。まず東京都 54区市町村を分析単位として、人口学的・社会経済的特性を示す7つの指標にもとづきクラスター分析を施し、標本抽出のために5つの層（地域類型）を析出した。次に、5つの層から各1地点ずつ確率比例によって調査対象地を選び、各地点から系統抽出法により20歳以上70歳未満の男女3000人を抽出した。具体的には、第1次抽出単位として、港区、大田区、世田谷区、清瀬市、あきる野市の5区市が選出され、各地点からそれぞれ600人が標本として抽出された。標本抽出にあたっては有権者名簿を用いたので、標本はすべて日本人である。調査は郵送法で行い、回答率は約33%であった。

この過程で生じるバイアスとして、次の3つが考えられる。

第1に、標本を抽出する際に生じる誤差（標本抽出誤差）である。5地点の標本は、たとえ回答率が100%であったとしても、5地点それぞれの母集団を正確に代表するものとはならず、確率論的に多少のバイアスが生じる。しかし、これは統計的検定によって考慮に入れることのできるバイアスであり、データ分析上は問題がない。

第2に、調査対象地として5地点を選択し、各地点に同数の標本を割り当てたことから生じるバイアスである。地点選択にあたっては、各類型に含まれる区市町村を人口によって重みづけしたうえで、無作為に抽出したので、個人レベルにおいては、理論上どの層（地域類型）に含まれる個人も等確率で当たる設計になっている。しかし、5地点の標本数は同数であるので、地区類型間の構成比は現実を反映したものではない。したがって5地点の標本を集計したものが、東京都全体を代表するものとは限らない。これも標本設計に起因するバイアスであるが、通常の統計的検定によっては考慮されないものである。

第3に、回答率の低さによるバイアスが考えら

れる。実際にわれわれが手にしている標本は、抽出された標本の3分の1にすぎない。3人に1人の回答がランダムに生じているという保証はなく、回答者の属性によって回答率が違っている可能性が高い。最も懸念されるのはこの種のバイアスである。

そこで本稿では、とくに第2、第3のバイアスに注目して、次の3つの水準で標本の代表性を検討する。①5地点それぞれについて、各地点の標本がどのくらい各地点の母集団を代表しているのか。②標本全体がどのくらい5地点を合計した母集団を代表しているのか。③標本全体がどのくらい東京都全体の母集団を代表しているのか（注(1)）。それぞれについて、バイアスがある場合には、どの層（回答者の属性）が標本抽出誤差を超えて過大／過小に代表されているかについても検討したい。

2. 方 法

ここでは、2000年10月——TGSSの直後——に実施された国勢調査のデータを母集団を示すものと想定し、現時点で公表されている東京都の性別・年齢別集計結果（日本人）をもとに、TGSS標本の代表性を検討する。

具体的には、次のような手順で分析を進めた。

①国勢調査結果をもとに、各地点における20歳以上70歳未満の日本人の性別・年齢別分布表、これを5地点全体について合算した性別・年齢別分布表、および東京都全体における対応する性別・年齢別分布表を作成した。

②これを母集団に関する統計表であるとみなして、性別構成、年齢構成、性別年齢構成の3つの次元で、母比率とTGSSの標本比率を比較した。性別構成の比較は、年齢階級ごとのサブサンプルについても行い、年齢構成の比較は、性別サブサンプルについても行った。いずれの場合にも、標本比率が、母比率にもとづいて計算された信頼区間を超えている場合には、標本抽出誤差を超えたバイアスがあると考えられる（注(2)）。各地点および5地点合計については、そのバイアスは回

答率に起因するものと推測できる。また、東京都全体との比較においては、5地点合計との乖離が大きいかぎりにおいて、調査設計上のバイアス（標本割り当てのバイアス）も加わっているものと解釈される。

③性別年齢別構成のクロス表については、標本が母集団を正確に反映している場合の期待値を基準としたカイ自乗検定も行った（注(3)）。この結果が有意な場合、クロス表全体として標本抽出誤差を超えたバイアスがあることが確認できる。

3. 性別構成比の検討

3. 1 各地点を母集団とみなした場合

まず、実際に調査の対象となった5区市それぞれにおける標本の代表性について、性別次元で検討しよう。表1は、5地点のそれぞれについて、5つのクロス表から構成されている。一番上のクロス表は、5地点それぞれにおける国勢調査の性別年齢別人口構成を実数で示したもので、これが分析のもとになる母集団の統計である。これを年齢層ごとの性別構成比に直したものが上から2番目のクロス表である。合計欄は各地区における母

表1 母集団における性別年齢別人口構成、年齢別性別構成比、および対応する標本比率、比率の差と有意水準
(港区、大田区、世田谷区)

平成12年国調	港区				大田区				世田谷区		
	男性	女性	計		男性	女性	計		男性	女性	計
20-29	11,456	12,121	23,577	20-29	56,623	52,181	108,804	20-29	77,185	76,904	154,089
30-39	12,365	12,973	25,338	30-39	53,732	46,442	100,174	30-39	66,356	67,647	134,003
40-49	9,393	10,273	19,666	40-49	43,089	38,206	81,295	40-49	48,857	49,339	98,196
50-59	11,407	12,699	24,106	50-59	52,354	48,493	100,847	50-59	51,784	54,813	106,597
60-69	8,023	10,001	18,024	60-69	35,281	39,033	74,314	60-69	38,472	44,569	83,041
計	52,644	58,067	110,711	計	241,079	224,355	465,434	計	282,654	293,272	575,926
性別構成比											
20-29	48.6%	51.4%		20-29	52.0%	48.0%		20-29	50.1%	49.9%	
30-39	48.8%	51.2%		30-39	53.6%	46.4%		30-39	49.5%	50.5%	
40-49	47.8%	52.2%		40-49	53.0%	47.0%		40-49	49.8%	50.2%	
50-59	47.3%	52.7%		50-59	51.9%	48.1%		50-59	48.6%	51.4%	
60-69	44.5%	55.5%		60-69	47.5%	52.5%		60-69	46.3%	53.7%	
計	47.6%	52.4%		計	51.8%	48.2%		計	49.1%	50.9%	
TGSS標本	男性	女性	n		男性	女性	n		男性	女性	n
20-29	40.0%	60.0%	20	20-29	50.0%	50.0%	22	20-29	33.3%	66.7%	18
30-39	51.5%	48.5%	33	30-39	43.6%	56.4%	39	30-39	48.7%	51.3%	39
40-49	53.1%	46.9%	32	40-49	50.0%	50.0%	26	40-49	39.4%	60.6%	33
50-59	54.3%	45.7%	46	50-59	47.5%	52.5%	40	50-59	47.8%	52.2%	46
60-69	53.2%	46.8%	47	60-69	43.4%	56.6%	53	60-69	38.5%	61.5%	39
計	51.7%	48.3%	178	計	46.1%	53.9%	180	計	42.9%	57.1%	175
百分率の差											
20-29	-8.6%	8.6%		20-29	-2.0%	2.0%		20-29	-16.8%	16.8%	
30-39	2.7%	-2.7%		30-39	-10.0%	10.0%		30-39	-0.8%	0.8%	
40-49	5.3%	-5.3%		40-49	-3.0%	3.0%		40-49	-10.4%	10.4%	
50-59	7.0%	-7.0%		50-59	-4.4%	4.4%		50-59	-0.8%	0.8%	
60-69	8.7%	-8.7%		60-69	-4.1%	4.1%		60-69	-7.8%	7.8%	
計	4.1%	-4.1%		計	-5.7%	5.7%		計	-6.2%	6.2%	
有意水準	男性	女性			男性	女性			男性	女性	
20-29				20-29				20-29			
30-39				30-39				30-39			
40-49				40-49				40-49			
50-59				50-59				50-59			
60-69				60-69				60-69			
計				計				計			

集団の性別構成比を示している。3番目のクロス表は、TGSS 標本における各地区の年齢別性別構成比を示したものであり、ここでも合計欄は各地区標本の性別構成比を示している。4番目の表は、対応するセルにおける標本比率 p と母比率 P の差 ($p-P$) を示したもので、正の値は標本比率が母比率を上回っていることを、負の値は標本比率が母比率を下回っていることを示している（合計欄についても同様）。男女を合計すればつねに100%であるから、男性が過小に代表されていれば、そのぶん女性は過大に代表されることになる。5番目のクロス表は、標本比率と母比率との差が、信頼区間 95% の範囲を超えているものについて記号で表示したものである。++は標本比率が信頼

区間 99% を超えて過大であることを、+は信頼区間 95% を超えて 99% 以下の範囲内で過大であることを、-は信頼区間 95% を超えて 99% 以下の範囲で過小であることを、そして--は信頼区間 99% を超えて過小であることを示している。換言すれば、++/-は危険率 1% 未満で、母比率との差が標本抽出誤差の範囲を超えて過大/過小であることを、+/-は危険率 5% 未満で、母比率との差が標本抽出誤差の範囲を超えて過大/過小であることをそれぞれ示している。この場合、母比率との差は標本抽出誤差によるものではなく、回答率の差によるものであることが推測されるわけである。

5 地点のうち、港区、大田区、世田谷区につ

表 1 (つづき)
(清瀬市、あきる野市)

平成12年国調	清瀬市				あきる野市		
	男性	女性	計		男性	女性	計
20-29	5,014	5,110	10,124	20-29	5,817	5,346	11,163
30-39	4,837	4,367	9,204	30-39	5,194	4,737	9,931
40-49	4,057	4,067	8,124	40-49	4,873	4,842	9,715
50-59	4,923	5,392	10,315	50-59	6,514	6,453	12,967
60-69	4,604	4,785	9,389	60-69	4,653	4,359	9,012
計	23,435	23,721	47,156	計	27,051	25,737	52,788
性別構成比							
20-29	49.5%	50.5%		20-29	52.1%	47.9%	
30-39	52.6%	47.4%		30-39	52.3%	47.7%	
40-49	49.9%	50.1%		40-49	50.2%	49.8%	
50-59	47.7%	52.3%		50-59	50.2%	49.8%	
60-69	49.0%	51.0%		60-69	51.6%	48.4%	
計	49.7%	50.3%		計	51.2%	48.8%	
TGSS標本							
20-29	41.7%	58.3%	24	20-29	32.1%	67.9%	28
30-39	32.5%	67.5%	40	30-39	44.8%	55.2%	29
40-49	40.5%	59.5%	37	40-49	39.0%	61.0%	41
50-59	38.0%	62.0%	50	50-59	35.6%	64.4%	73
60-69	47.7%	52.3%	65	60-69	53.8%	46.2%	52
計	40.7%	59.3%	216	計	41.3%	58.7%	223
百分率の差							
20-29	-7.8%	7.8%		20-29	-20.0%	20.0%	
30-39	-20.1%	20.1%		30-39	-7.5%	7.5%	
40-49	-9.4%	9.4%		40-49	-11.2%	11.2%	
50-59	-9.7%	9.7%		50-59	-14.6%	14.6%	
60-69	-1.3%	1.3%		60-69	2.2%	-2.2%	
計	-9.0%	9.0%		計	-9.9%	9.9%	
有意水準							
20-29				20-29	-	+	
30-39	-	+		30-39			
40-49				40-49			
50-59				50-59	-	+	
60-69				60-69			
計	--	++		計	--	++	

いては、性別構成比に有意差はない。たとえば、港区の場合、母集団を示す国勢調査の20歳以上70歳未満の性別構成比は、男性47.6%、女性52.4%であった(N=110,711)。一方、TGSS標本では、男性51.7%、女性48.3%であった(n=178)。この差は4.1%であるが、これは標本抽出誤差の範囲内である。年齢階級別に見た場合には、nが小さくなるので、標本抽出誤差の範囲が広がることになり、標本比率のぶれは大きくなる。とくに港区の20歳代、大田区の30歳代、世田谷区の20歳代と40歳代などで標本誤差が大きい、いずれも標本抽出誤差の範囲内にあり、統計的検定において考慮に入れることのできる誤差にとどまっている。ただし、大田区と世田谷区では、すべての年齢層において男性よりも女性のほうが多めであることに注意が必要である。

これに対して、清瀬市とあきる野市の場合には、いずれも男性比率が有意に少なく、女性比率が有意に多い(有意水準の合計欄を参照)。清瀬市の場合、男性の母比率は49.7%、標本比率は40.7%で、その差は-9.0%。危険率1%水準の有意差がある。この差は、男性の回答率が相対的に低かったこと、とくに30歳代で男性の回答率が女性に比べて低かったことに起因するものと考えられる。同様に、あきる野市の場合にも、男性の母比率と標本比率との差は-9.9%であり、1%水準の有意差がある。年齢サブサンプルでは、20歳代と50歳代で男性の標本比率が有意に低く、とくにこの年齢層で男女の回答率に格差があったことが推測される。

3. 2 5地点を母集団とみなした場合

次に、5地点全体を母集団とみなした場合にどのようなバイアスがあるかを検討しよう。表2の左半分は、上から順に、5地点全体についての国勢調査による性別年齢階級別人口構成、年齢別性別構成比、およびTGSS標本の年齢別性別構成比、標本比率と母比率の差とその有意水準を示したものである。

5地点全体を合算すると、男性の比率は国勢調査データでは50.1% (女性比率49.9%、N=

1,252,015)、TGSS標本では44.2% (女性比率55.8%、n=972)で、5.9%の差で男性が過小に(女性が過大に)代表されている。nが大きいために、信頼区間は狭まり、この差は標本抽出誤差を危険率1%水準で超えている。つまり、男性の標本比率は、標本抽出誤差の範囲を超えて過小なのである。その要因を、年齢別のサブサンプルによって検討すると、概して若年層ほど比率の差が大きく、20歳代および50歳代では、男性の標本比率が5%水準で母比率を有意に下回っていることがわかる。すでに見たように、清瀬市およびあきる野市では男性標本が有意に少なく、また、大田区と世田谷区でも、男性標本が系統的に少なめであった。こうした効果が累積して、標本全体では男性が過小になっているのである。おそらく、男性(とくに20歳代と50歳代)の回答率が相対的に低かったことが、このような結果を招来したものである。

3. 3 東京都全体を母集団とみなした場合

TGSS標本は、東京都全体の母集団をどの程度代表しているのだろうか。表2の右半分は、上から順に、東京都全体の性別年齢別人口構成(国調データ)、年齢別性別構成比、そしてTGSSデータによる年齢別構成比、標本比率と母比率の差とその検定結果である。TGSSデータは同表の左半分と同じで、5地点のデータを合計したものである。

ここでまず母集団について、国勢調査データの5地点合計と東京都全体の性別構成比を比較してみると、東京都全体のほうが0.9%男性が多くなっている。年齢別には、とくに20歳代では東京都全体の男性比率が5地点合計よりも1.8%多い。しかしながら、この差はわずかなものであり、5地点全体の母集団と東京都全体の母集団は、性別構成比に関するかぎり、それほど大きな違いはないといつてよからう。

次に、TGSS標本を東京都全体の母集団と比較してみよう。東京都全体の母集団における男性比率は51.0% (N=8,526,615)、TGSSデータにおける男性比率は44.2% (n=972)であり、その差は

表2 5地点全体と東京都全体の性別年齢別人口構成、年齢階級別性別構成比、
および対応する標本比率、比率の差と有意水準

平成12年国調	5地点全体				東京都全体		
	男性	女性	計		男性	女性	計
20-29	156,095	151,662	307,757	20-29	1,079,624	978,640	2,058,264
30-39	142,484	136,166	278,650	30-39	962,143	872,336	1,834,479
40-49	110,269	106,727	216,996	40-49	758,597	712,429	1,471,026
50-59	126,982	127,850	254,832	50-59	889,881	889,335	1,779,216
60-69	91,033	102,747	193,780	60-69	661,992	721,638	1,383,630
計	626,863	625,152	1,252,015	計	4,352,237	4,174,378	8,526,615
性別構成比							
20-29	50.7%	49.3%		20-29	52.5%	47.5%	
30-39	51.1%	48.9%		30-39	52.4%	47.6%	
40-49	50.8%	49.2%		40-49	51.6%	48.4%	
50-59	49.8%	50.2%		50-59	50.0%	50.0%	
60-69	47.0%	53.0%		60-69	47.8%	52.2%	
計	50.1%	49.9%		計	51.0%	49.0%	
TGSS標本							
	男性	女性	n		男性	女性	n
20-29	39.3%	60.7%	112	20-29	39.3%	60.7%	112
30-39	43.9%	56.1%	180	30-39	43.9%	56.1%	180
40-49	43.8%	56.2%	169	40-49	43.8%	56.2%	169
50-59	43.5%	56.5%	255	50-59	43.5%	56.5%	255
60-69	47.7%	52.3%	256	60-69	47.7%	52.3%	256
計	44.2%	55.8%	972	計	44.2%	55.8%	972
百分率の差							
20-29	-11.4%	11.4%		20-29	-13.2%	13.2%	
30-39	-7.2%	7.2%		30-39	-8.5%	8.5%	
40-49	-7.0%	7.0%		40-49	-7.8%	7.8%	
50-59	-6.3%	6.3%		50-59	-6.5%	6.5%	
60-69	0.7%	-0.7%		60-69	-0.1%	0.1%	
計	-5.9%	5.9%		計	-6.8%	6.8%	
有意水準							
	男性	女性			男性	女性	
20-29	-	+		20-29	--	++	
30-39				30-39	-	+	
40-49				40-49	-	+	
50-59	-	+		50-59	-	+	
60-69				60-69			
計	--	++		計	--	++	

-6.8%、危険率1%未満で有意差がある。すなわち、東京都全体の母集団に対して男性の標本は過小に代表されている（必然的に、女性の標本は過大に代表されている）わけである。また、年齢別のサブサンプルを検討すると、60歳代を除くすべての年齢層において、男性が過小に代表されていることがわかる。このことは、前項で見たように、そもそも男性の回答率が女性に比べて低かったことに加えて、東京都全体の母集団の男性比率が、5地点合計の母集団のそれよりもいくらか大きいことに起因するものである。また、概して若年層ほどこの差が大きくなっていることにも留意したい。

このように、標本の性別構成に関しては、いく

らか男性が過小に、女性が過大に代表されていることがわかった。次に、視点を変えて、年齢構成の観点から標本のバイアスを検討することにしよう。

4. 年齢別構成比の検討

4. 1 各地点を母集団とみなした場合

表3の最上段は、5地点それぞれについて、表1（最上段）の国勢調査による性別年齢別人口構成を年齢別の構成比として表したものである。この表の合計欄は、男女を合わせた年齢別構成比である。以下、順に、TGSS標本の年齢別構成比、

表3 母集団における性別年齢構成比および対応する標本比率、比率の差と有意水準
(港区、大田区、世田谷区)

平成12年国調	港区				大田区				世田谷区		
	男性	女性	計		男性	女性	計		男性	女性	計
20-29	21.8%	20.9%	21.3%	20-29	23.5%	23.3%	23.4%	20-29	27.3%	26.2%	26.8%
30-39	23.5%	22.3%	22.9%	30-39	22.3%	20.7%	21.5%	30-39	23.5%	23.1%	23.3%
40-49	17.8%	17.7%	17.8%	40-49	17.9%	17.0%	17.5%	40-49	17.3%	16.8%	17.1%
50-59	21.7%	21.9%	21.8%	50-59	21.7%	21.6%	21.7%	50-59	18.3%	18.7%	18.5%
60-69	15.2%	17.2%	16.3%	60-69	14.6%	17.4%	16.0%	60-69	13.6%	15.2%	14.4%
N	52,644	58,067	110,711	N	241,079	224,355	465,434	N	282,654	293,272	575,926
TGSS標本	港区				大田区				世田谷区		
	男性	女性	計		男性	女性	計		男性	女性	計
20-29	8.7%	14.0%	11.2%	20-29	13.3%	11.3%	12.2%	20-29	8.0%	12.0%	10.3%
30-39	18.5%	18.6%	18.5%	30-39	20.5%	22.7%	21.7%	30-39	25.3%	20.0%	22.3%
40-49	18.5%	17.4%	18.0%	40-49	15.7%	13.4%	14.4%	40-49	17.3%	20.0%	18.9%
50-59	27.2%	24.4%	25.8%	50-59	22.9%	21.6%	22.2%	50-59	29.3%	24.0%	26.3%
60-69	27.2%	25.6%	26.4%	60-69	27.7%	30.9%	29.4%	60-69	20.0%	24.0%	22.3%
n	92	86	178	n	83	97	180	n	75	100	175
百分率の差	港区				大田区				世田谷区		
	男性	女性	計		男性	女性	計		男性	女性	計
20-29	-13.1%	-6.9%	-10.1%	20-29	-10.2%	-12.0%	-11.2%	20-29	-19.3%	-14.2%	-16.5%
30-39	-5.0%	-3.7%	-4.4%	30-39	-1.8%	2.0%	0.2%	30-39	1.8%	-3.1%	-1.0%
40-49	0.7%	-0.3%	0.2%	40-49	-2.2%	-3.6%	-3.1%	40-49	0.0%	3.2%	1.8%
50-59	5.5%	2.5%	4.0%	50-59	1.2%	0.0%	0.5%	50-59	11.0%	5.3%	7.8%
60-69	12.0%	8.4%	10.1%	60-69	13.1%	13.5%	13.4%	60-69	6.4%	8.8%	7.9%
有意水準	港区				大田区				世田谷区		
	男性	女性	計		男性	女性	計		男性	女性	計
20-29	--	--	--	20-29	-	--	--	20-29	--	--	--
30-39				30-39				30-39			
40-49				40-49				40-49			
50-59				50-59				50-59	+		++
60-69	++	+	++	60-69	++	++	++	60-69		+	++

表3 (つづき)
(清瀬市、あきる野市)

平成12年国調	清瀬市				あきる野市		
	男性	女性	計		男性	女性	計
20-29	21.4%	21.5%	21.5%	20-29	21.5%	20.8%	21.1%
30-39	20.6%	18.4%	19.5%	30-39	19.2%	18.4%	18.8%
40-49	17.3%	17.1%	17.2%	40-49	18.0%	18.8%	18.4%
50-59	21.0%	22.7%	21.9%	50-59	24.1%	25.1%	24.6%
60-69	19.6%	20.2%	19.9%	60-69	17.2%	16.9%	17.1%
N	23,435	23,721	47,156	N	27,051	25,737	52,788
TGSS標本	清瀬市				あきる野市		
	男性	女性	計		男性	女性	計
20-29	11.4%	10.9%	11.1%	20-29	9.8%	14.5%	12.6%
30-39	14.8%	21.1%	18.5%	30-39	14.1%	12.2%	13.0%
40-49	17.0%	17.2%	17.1%	40-49	17.4%	19.1%	18.4%
50-59	21.6%	24.2%	23.1%	50-59	28.3%	35.9%	32.7%
60-69	35.2%	26.6%	30.1%	60-69	30.4%	18.3%	23.3%
n	88	128	216	n	92	131	223
百分率の差	清瀬市				あきる野市		
	男性	女性	計		男性	女性	計
20-29	-10.0%	-10.6%	-10.4%	20-29	-11.7%	-6.3%	-8.5%
30-39	-5.8%	2.7%	-1.0%	30-39	-5.1%	-6.2%	-5.8%
40-49	-0.3%	0.1%	-0.1%	40-49	-0.6%	0.3%	0.0%
50-59	0.6%	1.5%	1.2%	50-59	4.2%	10.8%	8.1%
60-69	15.6%	6.4%	10.2%	60-69	13.2%	1.4%	6.2%
有意水準	清瀬市				あきる野市		
	男性	女性	計		男性	女性	計
20-29	-	--	--	20-29	--		--
30-39				30-39			-
40-49				40-49			
50-59				50-59		++	++
60-69	++		++	60-69	++		+

対応するセルにおける標本と母集団の比率の差、そしてその有意水準を示している（合計欄は、いずれも男女を合わせた年齢別構成比に関するものである）。地点ごとに微妙な違いがあるので、順に確認していこう。

港区の標本では、全体として20歳代が過小であり、60歳代が過大である。性別サブサンプルの年齢別構成比では、この傾向は男性に強く現れている。

大田区の標本でも、20歳代が過小であり、60歳代が過大である。この傾向は、男性サブサンプルにおいても、女性サブサンプルにおいても、同様である。

世田谷区の標本では、20歳代が過小であり、50歳代と60歳代が過大である。男性サブサンプルでは20歳代が過小、50歳代が過大、女性サブサンプルでは20歳代が過小、60歳代が過大である。

清瀬市の標本では、20歳代が過小、60歳代が過大であり、この傾向は男性サブサンプルで強く現

れている。

あきる野市の標本では、20歳代と30歳代が過小、50歳代と60歳代が過大。男性サブサンプルでは、20歳代が過小、60歳代が過大、女性サブサンプルでは50歳代が過大になっている。

このように各地点とも、年齢構成上、若年層が過小に代表されており、中高年層が過大に代表される傾向がある。このことから、各地点とも、若年層の回答率が低く、中高年層とくに60歳代の回答率が比較的良好であったことが推測される（この表現は、回答率が全体として低かったことを踏まえたものである）。

4. 2 5地点を母集団とみなした場合

同様に5地点合計について、表4（左半分）を作成した。5地点全体のTGSS標本の年齢別構成比は、母比率に比べて20歳代と30歳代で過小に、50歳代と60歳代で過大に代表されている。この傾向は、とくに男性サブサンプルに顕著であり、男

表4 性別年齢構成比および対応する標本比率、比率の差と有意水準
(5地点全体、東京都全体)

平成12年国調	5地点全体				東京都全体		
	男性	女性	計		男性	女性	計
20-29	24.9%	24.3%	24.6%	20-29	24.8%	23.4%	24.1%
30-39	22.7%	21.8%	22.3%	30-39	22.1%	20.9%	21.5%
40-49	17.6%	17.1%	17.3%	40-49	17.4%	17.1%	17.3%
50-59	20.3%	20.5%	20.4%	50-59	20.4%	21.3%	20.9%
60-69	14.5%	16.4%	15.5%	60-69	15.2%	17.3%	16.2%
N	626,863	625,152	1,252,015	N	4,352,237	4,174,378	8,526,615
TGSS標本							
	男性	女性	計		男性	女性	計
20-29	10.2%	12.5%	11.5%	20-29	10.2%	12.5%	11.5%
30-39	18.4%	18.6%	18.5%	30-39	18.4%	18.6%	18.5%
40-49	17.2%	17.5%	17.4%	40-49	17.2%	17.5%	17.4%
50-59	25.8%	26.6%	26.2%	50-59	25.8%	26.6%	26.2%
60-69	28.4%	24.7%	26.3%	60-69	28.4%	24.7%	26.3%
n	430	542	972	n	430	542	972
百分率の差							
	男性	女性	計		男性	女性	計
20-29	-14.7%	-11.8%	-13.1%	20-29	-14.6%	-10.9%	-12.6%
30-39	-4.3%	-3.2%	-3.8%	30-39	-3.7%	-2.3%	-3.0%
40-49	-0.4%	0.4%	0.1%	40-49	-0.2%	0.4%	0.1%
50-59	5.5%	6.1%	5.8%	50-59	5.4%	5.3%	5.3%
60-69	13.9%	8.3%	10.8%	60-69	13.2%	7.4%	10.1%
有意水準							
	男性	女性	計		男性	女性	計
20-29	--	--	--	20-29	--	--	--
30-39	-	-	-	30-39	-	-	-
40-49				40-49			
50-59	++	++	++	50-59	++	++	++
60-69	++	++	++	60-69	++	++	++

性の20歳代と30歳代は過小、50歳代と60歳代は過大である。女性の場合にも、20歳代が過小、50歳代と60歳代が過大である。

前項で見たように、各地点とも、標本比率は若年層が過小に中高年層が過大に代表される傾向にあり、その累積結果として、5地点全体についても、母集団に比べて若年層が少なく中高年層が多くなる傾向がはっきり示されている。

4. 3 東京都全体を母集団とみなした場合

表4の右半分は、東京都を母集団とみなした場合の検討結果である。5地点合計の母集団と東京都全体の母集団の年齢構成は類似しているけれども、いくらか東京都全体において若年層が少なく、中高年層が多い。

TGSS標本を東京都全体の母集団を代表するものとみなした場合、どのようなバイアスが生じるのであろうか。5地点を母集団とする場合とほとんど変わりはない。若年層は過小に、中高年層は過大に代表される。この傾向は男女のサブサンプルのどちらにおいても同様に見られる（とくに男性において顕著である）。このことは、東京都全体と5地点合計との微妙なずれ（反対方向に働く）以上に、年齢階層別の回答率の違いが標本にバイアスを持ち込んでいることを意味している。

以上のように、年齢次元におけるTGSSデータのバイアスは、各地点においても、5地点合計および東京都全体を母集団とみなした場合でも、一貫して、若年層を過小に代表し、中高年層を過大に代表するものになっていた。この傾向は、標本抽出誤差を超えたものであり、その原因は若年層の回答率が相対的に低く高齢層のそれが比較的良好であったためと推測される。

5. 性別・年齢別構成比の検討

以上の分析では、性別と年齢の2つの次元に分けて、それぞれ独立に母比率と標本比率を比較してきた。すなわち、性別構成比の分析においては、全体および各年齢サブサンプルにおける性別構成比だけを問題にし、年齢構成比の分析においては、

全体および各性別サブサンプルにおける年齢構成比だけを問題にしてきた。性別次元においては、いくらか男性標本が多く、女性標本が少ないというバイアスがあることが明らかとなり、年齢次元においては、若年層が少なく、中高年層が多いというかなり顕著なバイアスがあることがわかった。最後に、2つの次元を組み合わせることができるセルによって定義される特定のサブサンプル（たとえば20歳代の男性）の標本比率が、それぞれ各地点、5地点全体、東京都全体における当該集団の母比率に比べて、有意に多いのか（少ないのか）どうかを検討しよう。

5. 1 各地点を母集団とみなした場合

5地点それぞれを母集団とみなした場合、性別年齢別の構成比は表5の1番上のクロス表のようになる。ここでは、各セルの比率は、各セルの度数を各地点の人口で割ったもの（ $P_{ij}=N_{ij}/N$ ）となる。以下、順に、TGSS標本における対応する比率（ $p_{ij}=n_{ij}/n$ ）、各セルの標本比率と母比率との差（ $p_{ij}-P_{ij}$ ）、そしてこの差が標本抽出誤差を超えたものであるかどうかを示した。たとえば、港区の20歳代の男性は、母比率10.3%（ $N=110,711$ ）、標本比率4.5%（ $n=178$ ）、比率の差が-5.8%で、5%水準の有意差がある。また、あわせてクロス表全体の適合度をカイ自乗検定によって示した。

港区の場合、標本抽出誤差を超えたバイアスは、20歳代の男性で過小、60歳代の男性で過大である。また、クロス表全体としても有意差がある。

大田区では、20歳代の男女が過小、60歳代の男女が過大であり、クロス表全体としても有意差がある。

世田谷区では、20歳代の男女が過小、60歳代の女性が過大である。クロス表全体としても有意差がある。

清瀬市では、20歳代の男女と30歳代の男性が過小、60歳代の男女が過大である。クロス表全体としても有意差がある。

あきる野市では、20歳代と30歳代の男性が過小、50歳代の女性と60歳代の男性が過大である。クロス表全体としても有意差がある。

表5 母集団における性別×年齢の構成比および対応する標本比率、比率の差と有意水準
(港区、大田区、世田谷区)

	港区			大田区			世田谷区	
平成12年国調	男性	女性		男性	女性		男性	女性
20-29	10.3%	10.9%	20-29	12.2%	11.2%	20-29	13.4%	13.4%
30-39	11.2%	11.7%	30-39	11.5%	10.0%	30-39	11.5%	11.7%
40-49	8.5%	9.3%	40-49	9.3%	8.2%	40-49	8.5%	8.6%
50-59	10.3%	11.5%	50-59	11.2%	10.4%	50-59	9.0%	9.5%
60-69	7.2%	9.0%	60-69	7.6%	8.4%	60-69	6.7%	7.7%
N	110,711		N	465,434		N	575,926	
TGSS 標本	男性	女性		男性	女性		男性	女性
20-29	4.5%	6.7%	20-29	6.1%	6.1%	20-29	3.4%	6.9%
30-39	9.6%	9.0%	30-39	9.4%	12.2%	30-39	10.9%	11.4%
40-49	9.6%	8.4%	40-49	7.2%	7.2%	40-49	7.4%	11.4%
50-59	14.0%	11.8%	50-59	10.6%	11.7%	50-59	12.6%	13.7%
60-69	14.0%	12.4%	60-69	12.8%	16.7%	60-69	8.6%	13.7%
n	178		n	180		n	175	
百分率の差	男性	女性		男性	女性		男性	女性
20-29	-5.8%	-4.2%	20-29	-6.1%	-5.1%	20-29	-10.0%	-6.5%
30-39	-1.6%	-2.7%	30-39	-2.1%	2.2%	30-39	-0.6%	-0.3%
40-49	1.1%	-0.9%	40-49	-2.1%	-1.0%	40-49	-1.1%	2.8%
50-59	3.7%	0.3%	50-59	-0.6%	1.3%	50-59	3.6%	4.2%
60-69	6.8%	3.4%	60-69	5.2%	8.3%	60-69	1.9%	6.0%
有意水準								
20-29	-		20-29	-	-	20-29	--	-
30-39			30-39			30-39		
40-49			40-49			40-49		
50-59			50-59			50-59		
60-69	++		60-69	++	++	60-69		++
χ^2 自乗値	26.66386		有意確率	0.00002		χ^2 自乗値	33.69167	
						有意確率	0.00000	
χ^2 自乗値	35.24817		有意確率	0.00000		χ^2 自乗値	35.24817	
						有意確率	0.00000	

表5 (つづき)
(清瀬市、あきる野市)

	清瀬市			あきる野市	
平成12年国調	男性	女性		男性	女性
20-29	10.6%	10.8%	20-29	11.0%	10.1%
30-39	10.3%	9.3%	30-39	9.8%	9.0%
40-49	8.6%	8.6%	40-49	9.2%	9.2%
50-59	10.4%	11.4%	50-59	12.3%	12.2%
60-69	9.8%	10.1%	60-69	8.8%	8.3%
N	47,156		N	52,788	
TGSS 標本	男性	女性		男性	女性
20-29	4.6%	6.5%	20-29	4.0%	8.5%
30-39	6.0%	12.5%	30-39	5.8%	7.2%
40-49	6.9%	10.2%	40-49	7.2%	11.2%
50-59	8.8%	14.4%	50-59	11.7%	21.1%
60-69	14.4%	15.7%	60-69	12.6%	10.8%
n	216		n	223	
百分率の差	男性	女性		男性	女性
20-29	-6.0%	-4.3%	20-29	-7.0%	-1.6%
30-39	-4.3%	3.2%	30-39	-4.0%	-1.8%
40-49	-1.7%	1.6%	40-49	-2.0%	2.0%
50-59	-1.6%	3.0%	50-59	-0.6%	8.9%
60-69	4.6%	5.6%	60-69	3.8%	2.5%
有意水準					
20-29	--	-	20-29	--	
30-39	-		30-39	-	
40-49			40-49		
50-59			50-59		++
60-69	+	++	60-69	+	
χ^2 自乗値	32.11819		有意確率	0.00000	
			χ^2 自乗値	36.53161	
			有意確率	0.00000	

このように、地点によって微妙な違いはあるものの、概して若年層、とくに若年男性で回答率が低く、標本が過小に代表される傾向にあり、中高年層、とくに中高年女性で回答率が比較的良好で、標本が過大に代表される結果になったものと思われる。

5. 2 5地点を母集団とみなした場合

5地点全体を合算したものが、表6の左半分である。標本比率が、標本抽出誤差を超えて過小であるのは、20歳代の男女と30歳代の男性、逆に過大であるのは、50歳代の女性と60歳代の男女である。クロス表全体としても有意な差がある。ここでも若年層、とくに若年男性の標本が過小で、中高年層、とくに中高年女性の標本が過大になっていることがわかる。

5. 3 東京都全体を母集団とみなした場合

表6の右半分は、東京都全体を母集団とみなし

た場合の計算結果である。ここでも、5地点全体を母集団とみなした場合と同様の結果となっている。クロス表全体としても有意であり、標本抽出誤差を超えたバイアスがある（5地点を母集団とした場合よりもカイ自乗値がいくらか小さくなっているのは、東京都全体の母集団のほうが、いくぶん若年層が少ないからである）。

6. 要約と結論

6. 1 要約

本稿では、性別、年齢、性別×年齢の3つの観点から、各地点、5地点合計、および東京都全体を母集団とみなした場合の標本の代表性について検討してきた。主な知見を母集団ごとに要約すると次のようになる。

港区の場合、性別構成は標本抽出誤差の範囲内にあった。しかし年齢構成では、若年層が過小に、

表6 性別×年齢の構成比と対応する標本比率、比率の差と有意水準

(5地点全体、東京都全体)

	5 地点全体			東京都全体			
平成 12 年国調	男性	女性		男性	女性		
20-29	12. 5%	12. 1%	20-29	12. 7%	11. 5%		
30-39	11. 4%	10. 9%	30-39	11. 3%	10. 2%		
40-49	8. 8%	8. 5%	40-49	8. 9%	8. 4%		
50-59	10. 1%	10. 2%	50-59	10. 4%	10. 4%		
60-69	7. 3%	8. 2%	60-69	7. 8%	8. 5%		
N	1, 252, 015		N	8, 526, 615			
TGSS 標本	男性	女性		男性	女性		
20-29	4. 5%	7. 0%	20-29	4. 5%	7. 0%		
30-39	8. 1%	10. 4%	30-39	8. 1%	10. 4%		
40-49	7. 6%	9. 8%	40-49	7. 6%	9. 8%		
50-59	11. 4%	14. 8%	50-59	11. 4%	14. 8%		
60-69	12. 6%	13. 8%	60-69	12. 6%	13. 8%		
n	972		n	972			
百分率の差	男性	女性		男性	女性		
20-29	-8. 0%	-5. 1%	20-29	-8. 2%	-4. 5%		
30-39	-3. 3%	-0. 5%	30-39	-3. 2%	0. 2%		
40-49	-1. 2%	1. 3%	40-49	-1. 3%	1. 4%		
50-59	1. 3%	4. 6%	50-59	1. 0%	4. 4%		
60-69	5. 3%	5. 6%	60-69	4. 8%	5. 3%		
有意水準	男性	女性		男性	女性		
20-29	--	--	20-29	--	--		
30-39	--	--	30-39	--	--		
40-49			40-49				
50-59		++	50-59		++		
60-69	++	++	60-69	++	++		
χ 自乗値	178. 65701	有意確率	0. 00000	χ 自乗値	160. 61006	有意確率	0. 00000

高齢層が過大に代表されていた。とくにこの傾向は、男性サブサンプルに目立っており、男性の回答率に年齢による差が大きかったことがうかがわれる。性別×年齢の観点からも、男性若年層で過小に、男性高齢層で過大に代表されていた。

大田区の場合も、性別構成は標本抽出誤差の範囲内にあり、年齢構成では、若年層が過小に、中高年層が過大に代表されていた。性別×年齢の観点からも、この傾向は再現された。性別にかかわらず、若年層で過小、高齢層で過大であった。

世田谷区の場合もまた、性別構成は標本抽出誤差の範囲内にあり、年齢構成では、若年層が過小に、中高年層が過大に代表されていた。しかし、性別サブサンプル別に年齢構成を見ると、男性のなかでは若年層が過小に、中高年層（50歳代）がやや過大になっており、女性のなかでは若年層が過小に、高齢層（60歳代）が過大になっていた。性別×年齢の観点からは、若年層の男女が過小に、60歳代の女性が過大に代表されていた。世田谷区の特徴は、50歳代男性の回答率が比較的良好であったことである。

清瀬市の場合、性別構成では、男性が過小に、女性が過大に代表されていた。この性別構成のバイアスは、30歳代に顕著であった。年齢構成の観点からは、若年層が過小に、高齢層が過大に代表されており、性別サブサンプル別には、男性では若年層が過小、高齢層が過大、女性では、若年層が過小であった。性別×年齢の観点では、20歳代の男女と30歳代の男性が過小に、60歳代の男女が過大に代表されていた。清瀬市の特徴は、とくに30歳代で男女の回答率の格差が大きく、年齢による回答率の差が男性に顕著であったことである。

あきる野市の場合も、性別構成では、男性が過小に、女性が過大に代表されていた。年齢サブサンプル別には、20歳代と50歳代でこの傾向が有意であった。年齢構成では、20歳代と30歳代が過小、50歳代と60歳代が過大であったが、性別サブサンプルの検討から、男性では20歳代が過小、60歳代が過大、女性では50歳代が過大であった。性別×年齢の観点からは、20歳代と30

歳代の男性が過小、50歳代の女性と60歳代の男性が過大であった。したがって、あきる野市の特徴は、50歳代の女性の回答率が比較的良好で、年齢による回答率の差が男性に顕著であったことにある。

以上が地域別にみた標本バイアスの要約である。次に5地点全体を集計して、5地点または東京都全体を代表させる場合について要約しよう。

5地点全体を母集団とみなす場合、性別構成では男性が過小に、女性が過大に代表されている。年齢サブサンプル別では、20歳代と50歳代でこの傾向が有意である。つまり、20歳代の男性の回答率は目立って低く、50歳代の女性の回答率は比較的良好である。東京都全体の場合、このバイアスは強化され、60歳代を除くすべての年齢層で、男性が過小である。これは、母集団において東京都全体の男性比率が5地点よりもいくぶん高いことによる。5地点全体の年齢別構成については、20歳代と30歳代が過小に、50歳代と60歳代が過大に代表されていた。このバイアスはリニアなもので、全体として年齢が高くなるにつれて回答率が高くなっていたことがうかがわれる。この傾向は男性サブサンプルにおいても、女性サブサンプルにおいても見られるが、とくに男性において目立っている。東京都全体においてもこのバイアスは変わりがないが、いくぶん弱まっている。それは、母集団において、東京都全体のほうが5地点よりも若年層がわずかに少なく、高齢層がわずかに多いからである。年齢×性別の観点からは、5地点全体を母集団とみなした場合と東京都全体を母集団とみなした場合で、結果は変わらない。どちらの場合でも、20歳代の男女と30歳代の男性が過小で、50歳代の女性と60歳代の男女が過大である。

以上から、TGSS 標本には、地点別にも全体としても、若干の標本バイアスがあることは否定できない。このバイアスは、年齢次元においては、概して男性が過小に、女性が過大に代表されており、その要因は女性よりも男性の回答率が低かったためであると考えられる。また、年齢次元においては、概して、若年層が過小に、高齢層が過大

に代表されており、その要因もまた年齢が若ければ若いほど回答率が低かったためと考えられる。そして年齢によるバイアスは女性よりも男性に強く表れていた。ただし、5地点全体と東京都全体とのずれはわずかなものであり、TGSS 標本を5地点を代表するものとみなす場合のバイアスと、東京都全体を代表するものとみなす場合のバイアスは、ほとんど同じである。

6. 2 評価と対策

性別・年齢別の回答率の差に起因する標本バイアスは、他の調査においても、同じように報告されている（注(4)）。また、経験知としてもよく知られており、驚くべきことではない。したがって、本稿によって明らかになった類のバイアスは、この調査に特異なものとは考えられない。しかし、このことは何の気休めにもならない。また、今回は検討できなかった社会経済的地位（職業、学歴など）次元でのバイアスも考えられる（経験的には、社会経済的地位が高いほうが回答率が高いと想定される）。こうしたバイアスにどう対処すべきであろうか。

ひとつの方法は、バイアスを修正すべく重みづけをすることである。もうひとつの方法は、重みづけをせずに性別・年齢・社会経済的地位などを統制したクロスセクションの分析に徹することである。どちらをとるかはその分析の目的による。われわれがもし、世論調査のように、都民全体の意見分布を知りたいのであれば、重みづけによる補正が必要であろう（たとえば、都民の何割が外国人の居住を歓迎しているかを知りたい場合）。ただし、重みづけは、データを人工的に補正する点で、危険が伴う。たとえば、若年層のサンプルを2倍にカウントすることによって年齢バイアスが補正できたとしても、若年層の回答者には高学歴者が多く、高学歴者がそもそも過大に代表されているとすれば、年齢補正のための重みづけは、学歴分布をさらに歪めてしまう結果となる。未知のバイアスがあることを想定すると、重みづけは危険であると言わざるを得ない（また、特定のサンプルの複製をつくることによって、一般に変数間の関

連を強めてしまう恐れがある）。われわれがもし、どういうタイプの人がどういう行動をとり、どういう意見を表明しがちであるかということを知りたいのであれば、重みづけをせずに、年齢・性別などの基本属性を統計的にコントロールしたクロスセクションの分析に徹するのが良策であろう。

本調査も含めて、多くの社会調査の目的は、複数の変数間の相関関係から因果関係を推定することにある（注(5)）。したがって、重みづけをせずに、クロスセクションの分析に徹する方法が安全であると思われる。ただし、その場合でも、若年男性の標本数が絶対的に不足しているというような標本数の限界は、いかんともしがたいことに留意すべきである。

注

- 1) このほかに、各地点がどれだけ層化された地域類型を代表しているかという問題が考えられる。各類型に属する地域を母集団とみなす分析によってこの問題を検討すること自体は可能である。しかし、本稿ではこの問題を追求しなかった。それは、かりに5地点のそれぞれが、性別年齢構成の点で各地域類型を適切に代表していたとしても、各類型から1地点ずつしか選択していないために、TGSS データの分析のなかで発見された地域特性が、典型的な特性であるのか、その地点に固有の特性であるのか、操作的には判断できないからである。現在の調査設計のもとでは、それは理論的にのみ判断できる問題である。理論的な判断がどうであろうと、分析上、地域特性は、年齢・性別などの個人属性を統制した後に残る地域差として検出されるものであるから、地域特性の検出にあたって人口構成上のゆがみはほとんど問題にならない。したがって、各地点が、人口構成上、抽出のもとになった層をどの程度代表しているかを検討することは、不必要であり、かえって誤解を招くものであると考えた。
- 2) 本稿では、標本を単純無作為抽出した場合に $t = \frac{(p-P)}{\sqrt{P(1-P)/n}}$ が標準正規分布することを利用して、 p が95%信頼区間と99%信頼区間を超えているかどうかを検討した（白倉 1983, p.68を参考にした）。ここで p は標本比率、 P は母比率、 n は標本数である。厳密には、本調査の標本は単純無作為抽出ではないので、計算された信頼区間（あるいは有意確率）は判断の目安にすぎない。しか

し、実際のデータ分析における統計的検定もまた、事実上、同じ前提に立って行われている。

- 3) $\chi^2_G = \sum \sum (n_{ij} - F_{ij})^2 / F_{ij}$ ここで、 n_{ij} は観測値、 F_{ij} は期待値で nP_{ij} 、ただし P_{ij} は該当するセルの母比率 N_{ij}/N 、 N は母集団の総数、 n は標本総数。この方式は、中尾啓子氏のご教示による。
- 4) たとえば、松本 (1999)、森岡編 (2000)、森岡・星 (2001) など、ほぼ同様の傾向が指摘されている。
- 5) ただし、将来、時系列的な比較をする場合に厄介な問題にぶつかる。社会構造や社会意識の変容は、変数間の関連が不変のまま、人口構成が変化するために、集合的なレベルで生じるのだろうか、それとも、変数間の関連が変わることによって生じるのだろうか (たとえば、学歴と意識との対応関係が変化するなど)。あるいはその両者が同時に起こるのかもしれない。そしてそのような変容には、回答率の変化も含まれているかもしれない。これらはやがて取り組まなければならない、理論的・

方法論的問題である。

参 考 文 献

- 白倉幸男 1983『標本設計の方法』直井優編著『社会調査の基礎』サイエンス社。
- 松本康 1999「年齢と社会構造——95年名古屋調査データによる分析」『名古屋大学社会学論集』20、45-80頁。
- 松本康・原田謙 2001「2000年東京版総合社会調査の概要——調査設計、調査方法、回答率」『総合都市研究』76号、17-23頁。
- 森岡清志編 2000『都市社会のパーソナルネットワーク』東京大学出版会。
- 森岡清志・星教士 2001「研究目的と調査概要」『年賀状分析による拡大パーソナルネットワークの研究』(文部省科学研究費補助金基盤研究(B)(1)研究成果報告書)。

Key Words (キー・ワード)

Representativeness of the Sample (標本の代表性), Sampling Error (標本抽出誤差), Response Rate (回答率)

Representativeness of the Sample of the General Social Survey in Tokyo 2000

Yasushi Matsumoto*

*Graduate School of Urban Science, Tokyo Metropolitan University
Comprehensive Urban Studies, No.78, 2002, pp.131-145

Representativeness of the sample of the General Social Survey in Tokyo we conducted in September 2000 is examined through comparison with the Population Census data of October 2000. In terms of gender, age, and gender \times age, how the sample represents the population of each of the five study areas, the areas in total, and the area of the Tokyo Metropolis as a whole is analyzed respectively.

Findings: For gender, the rates of the male group are generally underrepresented, the female overrepresented, beyond the sampling error, probably because the response rates of the male groups are considered to be lower than the female groups; for age, the rates of the younger groups are generally underrepresented, the elder overrepresented, probably because the younger the age, the lower the response rates of the age groups; for gender \times age, the bias on age is more apparent in the male than the female groups. Yet the difference in the gender-age composition between the five areas total and the Tokyo as a whole is so little that when the sample is regarded as a representative of the Tokyo as a whole, the bias is almost same as when it is regarded as that of five areas total. The detected biases are not specific to our Survey, but common to other recent surveys.

It is suggested that the most proper way to correct the biases is not to weight the data but to concentrate on crosssectional analyses, partly because the purpose of the survey is not to describe the distribution of opinions but to analyze the correlation between variables, and partly because a weighting procedure might intensify unknown biases on other dimensions of the sample such as socio-economic status.